Evolutionary and Swarm Computing for scaling up the Semantic Web

Gueret, C.D.M.; Schlobach, K.S.; Dentler, K.; Schut, M.C.

published in
Proceedings of the 24th Benelux Conference on Artificial Intelligence (BNAIC 2012)
2012

Evolutionary and Swarm Computing for scaling up the Semantic Web

Christophe Guéret Stefan Schlobach Kathrin Dentler Martijn Schut

VU University Amsterdam
De Boelelaan 1105, 1081HV Amsterdam

Abstract

The success of the Semantic Web, with the ever increasing publication of machine readable semantically rich data on the Web, has started to create serious problems as the scale and complexity of information outgrows the current methods in use, which are mostly based on database technology, expressive knowledge representation formalism and high-performance computing.

We argue that methods from computational intelligence (CI) can play an important role in solving these problems. In this paper we introduce and systemically discuss the typical application problems on the Semantic Web and discuss CI alternative to address the limitations of their underlying reasoning tasks consistently with respect to the increasing size, dynamicity and complexity of the data. Finally, we discuss two case studies in which we successfully applied soft computing methods to two of the main reasoning tasks; an evolutionary approach to querying, and a swarm algorithm for entailment.

This short paper is a summary of Guéret, C.; Schlobach, S.; Dentler, K.; Schut, M.; Eiben, G. "Evolutionary and Swarm Computing for the Semantic Web". IEEE Computational Intelligence Magazine, Special Issue on Semantic Web Meets Computational Intelligence, Vol. 7, No. 2 (May 2012)

1 Introduction

The World Wide Web is a decentralized system enabling the publication of documents and links between these documents on the Internet. A document is a piece of text, usually written in HTML and made available at a particular address (the URI). The links between documents are based on anchors put in these texts (hypertext links) and express a relation whose meaning depends on the interpretation made of the anchoring text. The Semantic Web uses the Web as a platform to publish and interlink data, rather than documents. This platform can then be used to build applications.

2 From logic proofs to optimisation problems

In their systematic analysis [3] Harmelen *et. al* argued that typical Semantic Web applications require a rather restricted set of basic reasoning tasks ("Entailment", "Consistency", "Mapping"). However, the Semantic Web combines data with Semantics, *i.e.* provide extra meaning to data, and thus yields problems related to data management as well. We therefore extend this analysis with a set of basic data manipulation tasks ("Querying", "Storage"). Table 1 contains a formal description of the different tasks Semantic Web applications have to perform, along with a short explanation of the traditional solving techniques.

Every algorithm currently being put to use on Semantic Web data has been designed as a logic based method operating over a finite set of curated triples T (a "knowledge base"). However, a typical triple set T on the Semantic Web is not static, nor curated, and all the guarantees of logical reasoning (completeness, soundness, determinism, ...) are lost. Instead, one can only *aim* at them and thus *optimize* towards these ideals. The consensual approach is to fit the Semantic Web data into a knowledge base by downloading, aggregating and curating subsets of its content. The problem is therefore adapted to fit the currently available solving methods, rather than being addressed with novel techniques. In order to find solving methods capable

Task	Formal definition	Traditional approach	Alternate formulation
Querying	Given T and a query Q, return the set of triples $\{t \in T\}$ such that $T \vdash t < Q$	Lookup and join	Constrained optimisation
Storage	Given T and a triple t return $T \cup t$	Centralized indices, Distributed Hash-tables	Clustering
Entailment	Given T. Derive $t \notin T$ with $T \vdash t$	Centralized and parallelized deduction (rules)	Multi-objective optimisation
Consistency	Given T. Check whether $T \vdash \perp$ (false)	Logical reasoning	Constrained optimisation
Mapping	Given T and a mapping condition c. Return $s, o \in T \times T$ such that $c(s, o)$ likely holds with respect to T	Similarities search between resources and classes. Inductive reasoning	Classification

Table 1: Tasks and traditional solving methods to make use of a set of triples T. In the table, $\vdash$ stands for logical entailment and $t < Q$ implies that t is an instance of Q.

of dealing with the complex character of the Semantic Web we propose to rephrase the logical formulations of the tasks as optimization problems (*c.f.* 4th column of Table 1).

Evolutionary and Swarm algorithms are known to perform well on optimization problems with large, and eventually dynamic, search spaces. We developed two use-cases leveraging these two family of algorithms for two of the tasks previously introduced: eRDF and Swarms.

eRDF [2] is an evolutionary computing based system solving "Querying" tasks. The queries are turned into a constraint satisfaction problem, which is in turn relaxed into a constrained optimisation. A population of answers is evolved until solutions that are good enough can be returned.

Swarms [1] proposes the usage of a swarm of micro-reasoners, each taking care of a part of a global set of entailment rules. The individuals of the swarm visit the graph and rewire it according to the knowledge they have.

3 Conclusion

The Semantic Web is a complex system made of large-scale, dynamic and potentially incoherent data. We showed how the typical tasks of Querying, Storage, Entailment, Consistency checking and Mapping can be rephrased from a logic problem into an optimization problem. This reformulation allows for considering evolutionary and swarm computing as a way to face the scaling and coherence challenges posed by the data from the Semantic Web.

References

[1] Kathrin Dentler, Christophe Guéret, and Stefan Schlobach. Semantic web reasoning by swarm intelligence. In *Proceedings of the 5th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2009) at ISWC*, October 2009.

[2] Eyal Oren, Christophe Guéret, and Stefan Schlobach. Anytime query answering in RDF through evolutionary algorithms. In *ISWC*, volume 5318, pages 98–113. Springer Berlin Heidelberg, 2008.

[3] Frank van Harmelen, Annette ten Teije, and Holger Wache. Knowledge engineering rediscovered: towards reasoning patterns for the semantic web. In *K-CAP*, pages 81–88, 2009.